Organizational Innovation & Computation in Evolutionary History

A Biological Perspective exploring 8 General Principles in complex computational systems with a brief introduction to models

David Krakauer, SFI.
Biological Organization as Structural Hierarchy
Allometric scaling laws are based on the idea that generic features of gases, such as the ideal gas law, can be understood by assuming atoms to be structureless 'billiard balls' undergoing elastic collisions. Despite these simplifications, the theory captures many essential features of gases and spectacularly predicts many of their coarse-grained properties. The original theory acted as a starting point for more sophisticated treatments incorporating detailed structure, inelasticity, quantum mechanical effects, etc., which allow more detailed calculations. Other examples include the quark model of elementary particles and the theories describing the evolution of the universe from the big bang. This approach has also been successful in biology, perhaps most notably in genetics. Again, the original Mendelian theory made simplifying assumptions, portraying each phenotypic trait as the expression of pairs of particles, each derived from a different parent, which assorted and combined at random in offspring. Nevertheless, this theory captured enough of the coarse-grained essence of the phenomena so that it not only provided the basis for the applied sciences of human genetics and plant and animal breeding, but also guided the successful search for the molecular genetic code and supplied the mechanistic underpinnings for the modern evolutionary synthesis. Although the shortcomings of these theories are well-recognized, they quantitatively explain an extraordinary body of data because they do indeed capture much of the essential behavior.

Scaling as a manifestation of underlying dynamics has been instrumental in gaining deeper insights into problems across the entire spectrum of science and technology, because scaling laws typically reflect underlying general features and principles that are independent of detailed structure, dynamics or other specific characteristics of the system, or of the particular models used to describe it. So, a challenge in biology is to understand the ubiquity of quarter-powers—to explain them in terms of unifying principles that determine how life is organized and the constraints under which it has evolved. Over the immense spectrum of life the same chemical constituents and reactions generate an enormous variety of forms, functions, and dynamical behaviors. All life functions by transforming energy from physical or chemical sources into organic molecules that are metabolized to build, maintain and reproduce complex, highly organized systems. We conjecture...
Evolution: Relationship Among Elements of Hierarchy
Evolution: Origin of Elements of Hierarchy
Organizational Principle 1: Duplication & Divergence

Key Reference:
Annual Review of Genetics
Vol. 38: 615-643 (Volume publication date December 2004)
DUPLICATION AND DIVERGENCE: The Evolution of New Genes and Old Ideas
Models...

Genome

Whole Genome Duplicated

P–P Interaction Network

Asymmetric divergence

Deletion of Links with Probability \( \delta = 1 - \gamma \)

Iteration

proteins x 2

links x 4
Models...

![Diagram of models](image)

A: Phase diagram of the limit degree distribution
B: Protein connectivity distribution
C: Average connectivity of first neighbor proteins

Wednesday, July 28, 2010
Microbial Net of Life
<table>
<thead>
<tr>
<th>Vertebrate Class</th>
<th>Immunoglobulin Isotype</th>
<th>Rearrangement, Hypermutation, Thymus and Spleen</th>
<th>Class Switch</th>
<th>Germinal Centre Formation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placental mammals</td>
<td>IgM, IgD, IgE, IgA</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Birds</td>
<td>IgM, IgA</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Amphibians</td>
<td>IgM, IgY, IgX</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Bony fish</td>
<td>IgM, IgD</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cartilaginous fish</td>
<td>IgM, IgW, IgNAF</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Jawless fish</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

No immunoglobulins

\( \xi(t) \)
Models...

Low Dimensions                      High Dimensions

**Innate immune system**

\[
\frac{dv_i}{dt} = v_i(r - sz) \\
\frac{dz}{dt} = k \sum_i v_i - dz
\]

**Adaptive immune system**

\[
\frac{dv_i}{dt} = v_i(r - \sum_j c_{ji} x_j - sz) \\
\frac{dx_i}{dt} = \sum_j k_{ji} v_i - \delta x_i \\
\frac{dz}{dt} = k \sum_i v_i - dz
\]
Organizational Principle 2: Horizontal Transfer

1. Packaging of DNA
2. Transfer
3. Entry/Acceptance
4. Splicing/Recombination
5. Stable Inheritance

Key Reference:

NATURE REVIEWS | MICROBIOLOGY
VOLUME 3 | SEPTEMBER 2005 | 675
HORIZONTAL GENE TRANSFER: PERSPECTIVES AT A CROSSROADS OF SCIENTIFIC DISCIPLINES
Operons: duplication and divergence or horizontal transfer?

Organizational Principle 2.1: Symbiosis & Symbiogenesis

Key References:
Konstantin Mereschkowsky. Symbiogenesis and the Origin of Species. 1926.

to the lines of our new curved ordinates. In like manner, the still
more bizarre outlines of other fishes of the same family of Chato-
dects will be found to correspond to very slight modifications of

Fig. 146. Argyropelecus affinis.

Fig. 147. Serranus diaphanous.

Fig. 148. Scaurus sp.

Fig. 149. Pimelodus.

Fig. 150. Polygrammus.

Fig. 151. Pseudotroptichus albus.

Similar co-ordinates; in other words, to small variations in the values
of the constants of the coaxial curves.

In Figs. 150-153 I have represented another series of Acantho-
pterygian fishes, not very distantly related to the foregoing. If we

Darcy W Thompson. On Growth and Form. 1917.
Chap XVII. “The Comparison of related forms”
Evolutionary invention and innovation. In general, genetic or phenotypic variation in organisms starts to accumulate in the center between two developmental barriers that are difficult for selection to overcome (constraint 1 and constraint 2) (A and C). Once selection overcomes these constraints and the variation becomes fixed in the population of a particular organism, it is called an "invention." (B) Sometimes an invention is of immediate value to the organism in its current environment and so can be defined as innovative. (D) Alternatively, an invention may become innovative only later, when there is a change in the environment. In both cases, the invention becomes an innovation because it forms the basis of a series of subsequent adaptive radiations. Invention need not imply innovation, which often depends on additional environmental events (D). Often inventions are recognized after the fact—sometimes long after they have emerged.

Albertson & Kocher, Heredity 2006.
Organizational Principle 3:
Innovations accelerate the production of Inventions

Key References:
Erwin, D. and Krakauer, D.C.
Mechanisms of Organizational Innovation. I. Neutral Networks
Organizational Principle 4: Functional degeneracy promotes exploration

Key References

M. Huynen, P. F. Stadler and W. Fontana
Smoothness within Ruggedness: The role of neutrality in adaptation,

J. P. Crutchfield, "When Evolution is Revolution---Origins of Innovation." In Evolutionary
Dynamics---Exploring the Interplay of Selection, Neutrality, Accident, and Function. J. P.
Crutchfield and P. Schuster, editors. Santa Fe Institute Series in the Science of Complexity.
Mechanisms of Organizational Innovation. II. Regulatory Dimension
Genotype:
Transcription Factors Expression

\[
p(D_{ij} = +1 | c_{ij} = 1) = q \quad \tilde{p} = H(D\tilde{g})
\]
Organizational Principle 5: Regulatory epistasis promotes Diversity & inhibits Disparity
(A scale dependent conformity effect)

Key Reference:
Evolution, Inference & Computation
"In the struggle for survival, the fittest win out at the expense of their rivals because they succeed in adapting themselves best to their environment."

"I have called this principle, by which each slight variation, if useful, is preserved, by the term Natural Selection"
\[
\frac{\Delta g_i(t)}{\Delta t} = g_i(t - 1)(r_i(\mathbf{g}) - \bar{f})
\]

"In the struggle for survival, the fittest win out at the expense of their rivals because they succeed in adapting themselves best to their environment."

"I have called this principle, by which each slight variation, if useful, is preserved, by the term Natural Selection"
\[ P(H|E) = P(H) \frac{P(E|H)}{P(E)} \]
\[
P(H|E) = P(H) \frac{P(E|H)}{P(E)}
\]

\[
P(H|E) = P(H) \frac{L_H}{\bar{L}}
\]

\[
P_X(t) = P_X(t-1) \frac{L_X}{\bar{L}}
\]

\[
\bar{L} = P(E) = \sum_{x \in \omega} P(E|H)P(H)
\]
\[ \frac{\Delta g_i(t)}{\Delta t} = g_i(t - 1)(r_i(g) - \bar{f}) \]

\[ \Delta P_X(t) = P_X(t - 1)(f_t - \bar{f}), \quad \text{where} \quad f_t = \frac{L_X}{\bar{L}} \]

Baye's Rule

Good Idea!

replicator equation

see: Cosma Shalizi CMU statistics & others...
“Biological offspring are the hypotheses of their parents, formulated through the success of their ancestors at predicting the future state of the world”
Organizational meta-Principle 6:
Natural Selection is a process of Competitive Inference
(biological evolution is a subclass)
Inventions and Innovations driving the Evolution of Biological Computation
The \textit{lac} Operon and its Control Elements

\begin{itemize}
\item \textit{lacI} \hspace{1cm} \textit{lacZ} \hspace{1cm} \textit{lacY} \hspace{1cm} \textit{lacA}
\end{itemize}

\begin{itemize}
\item CAP \hspace{1cm} P \hspace{1cm} AUG \hspace{1cm} AUG \hspace{1cm} AUG \hspace{1cm} AUG \hspace{1cm} 3'
\end{itemize}

\text{mRNA}

\begin{itemize}
\item cAMP Activator Protein \hspace{1cm} RNA Polymerase
\end{itemize}

\begin{itemize}
\item High (constitutive) level of expression
\end{itemize}

\begin{itemize}
\item Low glucose
\item Lactose available
\end{itemize}

\begin{itemize}
\item \textit{lacI} repressor
\end{itemize}

\begin{itemize}
\item High glucose
\item Lactose unavailable
\end{itemize}

\begin{itemize}
\item Low glucose
\item Lactose unavailable
\end{itemize}

\begin{itemize}
\item Low (basal) level of expression
\end{itemize}

\begin{itemize}
\item High glucose
\item Lactose available
\end{itemize}
Models and Theory
\[
p_i = \frac{[X_i]/K_i}{1 + [X_i]/K_i}
\]

\[
p_j = \frac{[X_j]/K_j}{1 + [X_j]/K_j}
\]

\[
P = \frac{Z_{ON}}{Z_{ON} + Z_{OFF}}
\]
\[ p_i = \frac{[X_i]/K_i}{1 + [X_i]/K_i} \]

\[ p_j = \frac{[X_j]/K_j}{1 + [X_j]/K_j} \]

\[ P = \frac{Z_{ON}}{Z_{ON} + Z_{OFF}} \]

\[ Z_{OFF} = \sum_{\sigma_1 \in \{0,1\}} \cdots \sum_{\sigma_n \in \{0,1\}} W[\sigma_1, \ldots, \sigma_n] \]

\[ Z_{ON} = \sum_{\sigma_1 \in \{0,1\}} \cdots \sum_{\sigma_n \in \{0,1\}} Q[\sigma_1, \ldots, \sigma_n].W[\sigma_1, \ldots, \sigma_n] \]

\[ W[\sigma_1, \ldots, \sigma_n] = \prod_{i=1}^{n} q_{\sigma_i} \prod_{i<j}^{n} \omega_{i,j}^{\sigma_i \sigma_j} \]

\[ Q = q_p \prod_{i}^{n} [1 - \sigma_i \delta(\omega_{0i}, 0)][1 + \omega \sum_{j}^{n} \sigma_j \delta(\omega_{pj}, 0)] \]

\[ \omega_{i,j} = \{0, 1, \omega, \Omega\} \]

Buchler et al. On Schemes of combinatorial transcription logic. PNAS. 100. 2003
Interaction energy to the OR and NAND gates (see Fig. 2). The AND gate can implement the responses for the genes of physiological TF concentrations (the full response characteristics additional cooperative interaction). This is quantitatively verified by but when both are present binding occurs with the help of the other (see Fig. 2). The AND gate is referred to as the AND gate. It can be obtained by choosing weak interaction patches (15). For a number of well studied proteins (see refs. 10, 20, and references therein), such interactions fall within the relevant range of cellular protein concentrations (e.g., the conservative value of 10,000 nM) individually for each site.

Given the binding strengths above, let us consider the response of the desired effects, the operator sites need to be strong (filled boxes) and the promoter needs to be weak (open box). In this and subsequent plots are insensitive to the precise values of the parameters used.

### Supporting Text

To illustrate how different regulatory interaction with proteins bound to two sites result if the respective binding sites overlap. No effective interaction (0). The task of implementing various regulatory functions is then reduced to arranging the binding sites in the cis-regulatory constructs, we use the offset, overlapping boxes to indicate mutual repression and the dashed lines to indicate cooperative interaction.

#### cis-regulatory constructs

The above example illustrates a fundamental difference in the architecture integrating the regulatory regions of E. coli lac and the promoters positioned sequentially in the regulatory region, with many potential problems associated with their expressions (e.g., between each pair of sites can be selected from the values \( \{0, 1, 2, 4, 12, 20\} \) k cal mol, we see that interaction...
Organizational Principle 7: Positive & Negative Regulation Principle

Negative regulation dominates in microbes where products and contexts are limited.
Positive regulation dominates in Eukaryotes where products and contexts are numerous
Organizational Principle 8: Common Computational Representation

Regulatory architectures in complex systems can often be represented formally as networks of simple logical gates. These can be connected to represent a complex system as a circuit.
Network Representations: From Logic to Statistics of Structure

(a) Basic unit
(b) Motifs
(c) Modules
(d) Transcriptional regulatory network

Transcription factor

Target gene and binding site

SIM
MIM
FFL

Current Opinion in Structural Biology

Eight “Principles” of Organization in Complex Computational Systems

- P1: duplication & divergence
- P2: lateral transfer / recombination
- P3: innovation through accelerated invention
- P4: functional degeneracy
- P5: regulatory epistasis
- P6: evolution as Inference
- P7: neg’ and pos’ Regulation
- P8: common computational representations